Evaluating the Finite Sample Properties of Baseline Covariate Adjustment in Randomized Trials: Application to Time to Event Outcomes

by

Su Jin Lim

A thesis submitted to The Johns Hopkins University in conformity with the requirements for the degree of Master of Science.

Baltimore, Maryland

April, 2018

Abstract

In two-arm randomized controlled trials (RCTs) with baseline covariates that are prognostic for the outcome of interest, baseline covariate adjustment can improve precision of the estimated marginal treatment effect and increase statistical power, for a fixed sample size. Many adjusted estimators for the marginal treatment effect have been proposed for a variety of outcomes and the statistical properties of these estimators have been demonstrated via simulation studies in large sample RCTs. However, there is little guidance on the use of these adjusted estimators in RCTs with small sample sizes.

Motivated by an ongoing RCT (TREAT Parents trial), we conduct a simulation study to evaluate the statistical behavior of adjusted estimators for both time to event and binary outcomes in small sample RCTs, using the adjusted estimator for the marginal log hazard ratio proposed by Lu an Tsiatis (2008) for the primary time to event endpoint and several adjusted estimators of the marginal risk difference for the secondary binary endpoint. We considered hypothetical trials with small effective sample sizes, i.e. the expected number of events, ranging from 20 to 100. We

also considered 3 scenarios with varying degrees of correlation between the baseline covariates and the primary and secondary outcomes, including the scenario where the baseline covariates are independent of the outcomes. The hypothetical trials were generated in ways that mimic the observed characteristics within TREAT Parents trial.

Our simulation results demonstrated that precision gains in the estimated marginal treatment effect can be achieved when adjusting for baseline variables that are correlated with the outcome at effective sample sizes greater than 30. However, the potential precision gains depend on the strength of the assumed correlation between the baseline covariates and the outcome. For small effective sample sizes, the potential loss in precision when using an adjusted estimator when the baseline covariates are uncorrelated with the outcome can be roughly the same size of the precision gain under an assumption of modest correlation. We demonstrate the use of a simple simulation setting that may be used by researchers interested in evaluating the use of an adjusted estimator in a small sample RCT.

Primary Reader: Elizabeth Colantuoni, PhD

Secondary Reader: Michael Rosenblum, PhD

Acknowledgments

I would like to express my gratitude to my thesis advisor Elizabeth Colantuoni, PhD for her guidance, encouragement and continuous support during my Sc.M study and thesis research, and thesis reader Michael Rosenblum, PhD for his valuable suggestions and comments during my thesis research. I would like to thank Aaron Milstone, MD for supporting this thesis research and sharing the data from the Treating Parents to Reduce NICU Transmission of *Staphylococcus aureus* (TREAT Parents) trial, funded by NIH/AHRQ [R01HS022872].

I thank my friends and fellow graduate students for their academic and emotional support and friendship. Finally, I would like to thank my family for the continuous and unconditional support, encouragement and love they have given me during my years of study. This work would not have been possible without them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

We consider two-arm randomized controlled trials (RCTs) where the goal is to assess the effectiveness of a treatment relative to a control or placebo condition. The goal is to estimate the marginal or average treatment effect; e.g. for a linear outcome, the marginal treatment effect is defined as the difference in the population mean outcome under treatment and the population mean outcome under placebo. In RCTs, extensive baseline covariates are collected on each subject prior to randomization of the treatment assignment, including demographics, medical and treatment history, current symptoms or diseases, and other quantitative health measurements [1]. Typically, these baseline covariates are used to describe the study population, can be compared across treatment arms to quantify chance imbalance in the baseline covariates after randomization [1, 2], and may be incorporated into the estimation of the marginal treatment effect [2], i.e. baseline covariate adjustment. The key benefit

of baseline covariate adjustment is to exploit chance imbalance in baseline covariates that are prognostic for the outcome of interest with the goal of improving the precision of the estimated marginal treatment effect. That is, when adjusting for prognostic baseline covariates, the standard error for the estimated marginal treatment effect is smaller than that obtained by the unadjusted estimate of the marginal treatment effect [2], resulting in a gain of statistical power for a fixed sample size when comparing the adjusted to unadjusted estimator [3].

Adjusted estimators have been proposed for a variety of outcomes, including outcomes that follow an exponential family distribution [4,5], time to event outcomes [6,7] and ordinal outcomes [8]. These proposed adjusted estimators all share the desired property that, asymptotically (i.e. as the sample size goes to infinity), the estimated marginal treatment effect based on the adjusted estimator is guaranteed to be as precise or more precise than the unadjusted estimator. Many of the estimators have been evaluated, via simulation studies [3, 9], for RCT sample sizes which are generally accepted as "large" in practice (e.g. regulatory Phase III trials, such as [10], or non-regulatory trials, such as [11]). These simulation studies demonstrate the potential gains in precision when adjusting for prognostic baseline covariates. For instance, using simulation studies based on data from the completed MISTIE phase II trial for a novel treatment for intraventricular hemorrhage, Colantuoni and Rosenblum (2015) [3] generated hypothetical phase III trials, each with sample size of 412, targeting the marginal difference in the risk of a success, defined as the difference

between the population proportion of patients with a modified Rankin Scale (mRS) score of 3 or less under the treatment vs. control arm [3, 11]. The simulation results showed that the relative efficiency (RE, i.e. the ratio of the variance of the unadjusted to adjusted estimators) ranged from roughly 1.2 to 1.3 when baseline covariates were prognostic for the outcomes, as observed in the phase II trial [3]. Note that $1 - \frac{1}{RE}$ translates roughly to the relative reduction in required sample size when using the adjusted estimator compared to the unadjusted estimator, for a fixed power. Therefore, the range of relative efficiency translate to a 17 to 23% reduction in required sample size. In addition, these simulation studies demonstrated the potential loss of precision that an adjusted estimator may experience when the selected baseline covariates are uncorrelated with the outcome. For example, in the same simulation study that uses the MISTIE II trial data, the precision losses ranged from 0.7% to 2.8% comparing the adjusted estimator to the unadjusted estimator [3].

The established asymptotic and practical "large" sample size behavior of the proposed adjusted estimators are essential for researchers planning RCTs with "large" sample sizes. However, there is little to no guidance on the use of the adjusted estimators for researchers designing and conducting "small" RCTs, e.g. single center non-regulatory trials or small Phase II regulatory trials. The goal of this paper is to demonstrate the potential limitations of baseline covariate adjustment in RCTs with "small" sample sizes and offer guidance on the use of these estimators in "small" sample RCTs.

We are motivated by an ongoing RCT of parent-neonate dyads within two Johns Hopkins Hospital affiliated Neonatal Intensive Care Units (NICU), complete details of the TREAT Parents trial in Section 3.1 [12]. The goal of the trial is to evaluate whether treating, with antibiotics vs. placebo, parent(s)/caregiver(s) who are colonized with *Staphylococcus aureus* (*Staph aureus*) at the time of admission of their neonate to the NICU reduces the risk of the neonate being colonized with the parental/caregiver strain of *Staph aureus*, which we will refer to as concordant colonization. The RCT was designed to have 80% power to detect a 60% reduction in the hazard of concordant colonization comparing the treated and placebo arms, requiring a total of 40 concordant colonization events to be observed during the course of the trial. Recall that this number of required events represents the effective sample size for the trial and thus this trial would be considered "small". Thus, we consider RCTs where the outcome of interest is the time of a defined event where observation is within a fixed follow-up period. This outcome can be evaluated either as a survival endpoint, where the marginal treatment effect is defined as the relative hazard of the event comparing the treatment to placebo arm, or as a binary indicator, where the marginal treatment effect is defined as the absolute difference in the risk of the event comparing the treatment to placebo arm.

We consider the baseline covariate adjusted estimator for the marginal hazard ratio proposed by Lu and Tsiatis (2008) [7] and several baseline covariate adjusted estimators of the marginal risk difference, see Colantuoni and Rosenblum (2015) [3].

In this paper, we demonstrate, via simulation studies designed using a blinded subset of data from the ongoing TREAT Parents trial, the potential gains and losses in precision when using the selected adjusted estimators compared to the unadjusted estimator and make recommendations for researchers planning "small" sample RCTs.

Chapter 2

Notation and Definitions

2.1 Notation

Within the two-arm randomized trial of N participants, each participant's data is a vector (W_i, A_i, T_i, Y_i). For the i^{th} participant, W_i is a vector of baseline covariates and A_i is the treatment arm indicator ($A_i = 1$ for treatment, and $A_i = 0$ for placebo). We define two outcomes; a survival outcome $T_i = min(T_i^*, C_i)$ which is the minimum of the time when the event occurs (T_i^*) and the censoring time (C_i), and a binary outcome Y_i, where $Y_i = 1$ if $T_i^* < C_i$, and $Y_i = 0$ if $T_i^* \geq C_i$. Throughout, we assume that W, T and Y are observed for all patients (i.e. no missing values for the baseline covariates or outcomes) and the treatment is assigned with allocation 1:1 (i.e. each patient has probability 1/2 to receive the treatment). As a result of the randomization, we assume that A is independent of W. We also assume that the

censoring time, C, is independent of (T^*, W) conditional on the treatment assignment A.

2.2 Time to Event Outcome

For the time to event outcome, we assume that the proportional hazards model holds; i.e. $\lambda_{T|A}(t|a) = \lambda(t)exp(\beta a)$, where $\lambda_{T|A}(t|a)$ is the conditional hazard of the event at time t given treatment $A = a$, for $a = 0, 1$. We define the marginal treatment effect as the log hazard ratio β, i.e. the log of the hazard of concordant colonization comparing the treatment and placebo arms. The unadjusted estimator for β, β_{PH} is the maximum partial likelihood estimator for the coefficient for the main term of treatment, from a Cox proportional hazards model that includes only the main term A. Under the proportional hazards model, then the Wald test for β_{PH} is asymptotically equivalent to the log-rank test for treatment differences [13].

Using semiparametric theory, Lu and Tsiatis (2008) [7] proposed an estimator for the marginal log hazard ratio that incorporates baseline covariates and is as precise or more precise than the maximum partial likelihood estimator, β_{PH}, under no additional assumptions than described above. They derive estimating equations whose solutions contain all semiparametric estimators for β, including the maximum partial likelihood estimator. The baseline covariate information is incorporated into both the estimation of the underlying cumulative hazard function, via the Breslow estimator, and the

estimation of the cumulative hazard function of the censoring distribution, via the treatment-specific Nelson-Aalen estimator.

We note that alternative marginal treatment effect definitions for time to event outcomes include the difference in the mean restricted survival time [6].

2.3 Binary Outcome

For the binary outcomes, we define the marginal treatment effect as the risk difference, the difference between the population proportion of neonates concordantly colonized in 90 days post-randomization in the treatment vs control arm. The unadjusted estimator for the marginal risk difference is the difference between the observed proportion of neonates who were concordantly colonized in 90 days in treatment versus control arm. The unadjusted estimator is consistent and asymptotically normal, but it ignores information in the baseline covariates W. We consider three adjusted estimators for the marginal risk difference for evaluating their performance relative to the unadjusted estimator: the model standardization estimator, doubly robust weighted least squares (DR-WLS) estimator, and the "precise, locally efficient, augmented, simple estimator" (PLEASE).

The model standardization estimator requires two logistic regression working models for $P(Y = 1|A = a, W)$, fitted separately for each treatment assignment, $a = 0, 1$. The fits of these two models are then applied to all N participants such that estimated

probabilities of $P(Y = 1|A = a, W)$ for $a = 0, 1$ are obtained assuming separately that all N participants received $A = 1$ and then all participants received $A = 0$. The estimate of the marginal risk in the treatment arm, i.e. $P(Y = 1|A = 1)$ is obtained by taking the average of the N estimated probabilities. The marginal risk in the placebo arm is estimated similarly and the marginal risk difference is the difference in the two estimated treatment arm specific marginal risks. Moore and van der Laan [14] showed that the model standardization estimator is consistent for the marginal risk difference even if the logistic regression working models are not correctly specified. The DR-WLS estimator [15] is similar to the model standardization estimator; the only difference is the creation of a logistic regression working model for the treatment assignment, i.e. $P(A = 1|W)$, also known as the propensity score model, which is used in the fitting of the logistic regression working models. For the treatment arm, the $P(Y = 1|A = 1, W)$ is modeled as a weighted logistic regression model with weights defined as 1 divided by the predicted values for $P(A = 1|W)$ from the propensity score model. For the placebo arm, the $P(Y = 1|A = 0, W)$ is modeled as a weighted logistic regression model with weights defined as 1 divided by 1 minus the predicted values for $P(A = 1|W)$ from the propensity score model. After fitting the weighted logistic regression models, the estimator is implemented in the same way as the model standardization estimator. The double-robustness property of the DR-WLS estimator guarantees consistency of this estimator as long as either the propensity score model or logistic regression working model is correctly specified.

Although both the model standardization and DR-WLS estimators are consistent for the marginal treatment effect; these estimators are not guaranteed to be as precise or more precise than the unadjusted estimator, as N goes to infinity. Therefore, we consider one additional adjusted estimator, the PLEASE, which achieves this property. The PLEASE is a special case of the general class of estimators proposed by Rotnitzky et al. [16] and starts by fitting the same weighted logistic regression working models as used in the DR-WLS estimator. Then, two additional covariates are computed, one for each treatment arm, that are defined as the difference between the predicted $P(Y = 1|A = a, W)$ and the estimated marginal risk, $\hat{P}(Y = 1|A = a)$. The propensity score model is then refit including these two additional covariates, the weighted logistic regression models are refit and the marginal risks per treatment are computed. These additional covariates that are included in the propensity score model allow this estimator to achieve the desired asymptotic property.

In our simulation studies, we evaluate all three of the adjusted estimators defined above for estimating the marginal risk difference.

Chapter 3

Simulation Study

3.1 The TREAT Parents Trial

The Treating Parents to Reduce Neonatal Transmission of *Staphylococcus aureus* (TREAT Parents) Trial is a multicenter randomized, masked, placebo-controlled trial of neonates receiving care in the Neonatal Intensive Care Unit (NICU) to evaluate the efficacy of treating parent(s)/caregiver(s) colonized with *Staphylococcus aureus* (*Staph aureus*) with short term intranasal mupirocin and topical chlorhexidine antisepsis to reduce the transmission of the parental or caregiver strain of *Staph aureus* to the neonate. The neonate-parent dyad, the unit of analysis, was randomized into the treatment arm (intranasal mupirocin and topical antisepsis with chlorhexidine cloths) or placebo arm (placebo intranasal ointment and placebo cloths for skin antisepsis). The primary endpoint is time to concordant colonization, i.e. neonatal acquisition of

an *Staph aureus* strain that is concordant to the parental or caregiver *Staph aureus* strain. The observation time was censored either administratively at 90 days after randomization, when a neonate died or was discharged from the NICU. The key secondary endpoint is the binary indicator of concordant colonization by 90 days post randomization.

With the assumed placebo arm concordant colonization rate of 10%, the trial was designed to achieve 80% power to detect a marginal log hazard ratio of -0.92 (i.e. a hazard ratio of 0.40 or a 60% reduction in the hazard of concordant colonization) in the treated arm relative to placebo, requiring 40 total concordant colonizations. The pre-planned statistical analysis of the primary endpoint included estimation of the marginal log hazard ratio via the method proposed by Lu and Tsiatis (2008) [7] with adjustment for three baseline variables collected at the time of NICU admission that were thought to be correlated with the primary outcome: birth weight in grams, participating hospital (Johns Hopkins Hospital (JHH) vs. Bayview Medical Center (BMC)) and an indicator for whether the neonate was born at the participating hospital (inborn) or transferred from home or another hospital (outborn).

3.2 Data

Blinded data on 150 neonate-parent dyads collected by June 16, 2017 were used for evaluation. Table 3.1 shows descriptive statistics for the three baseline variables of

the 150 neonate-parent dyads stratified by the binary concordant colonization status by 90 days of the study. Among the 150 neonates, 90 percent were inborn (n = 135) and 93 percent were recruited at the JHH site (n=139). The average birth weight is 2049 grams (standard deviation, sd=978). A total of 25 neonates were concordantly colonized by 90 days. Neonates who were concordantly colonized by 90 days had, on average, 175g lower birth weight than those who did not become concordantly colonized, but the difference between the means was not statistically significant (p-value = 0.4).

Table 3.1: Descriptive statistics for three baseline variables from 150 neonates in the TREAT Parents trial. Values represent n(%) for birth location and admitted hospital and mean(sd) for birth weight.

Baseline Variable	Concordant Colonization (n=25)	No Concordant Colonization (n=125)	All neonates (n=150)
Birth Location: Inborn	23 (92)	112 (89.6)	135 (90)
Admitted Hospital: JHH	22 (88)	117 (93.6)	139 (92.7)
Birth Weight (g)	1904 (871)	2079 (999)	2049 (978)

To quantify the correlation between the three baseline covariates and concordant colonization, we fit a Cox proportional hazards model that included main terms for each of the three baseline covariates. The estimated coefficient for the main term assumes proportional hazards conditional on the baseline covariates. Table 3.2 displays the estimated hazard ratios from this model. After conditioning on birth weight and admitted hospital, we estimate that the hazard of concordant colonization is 35%

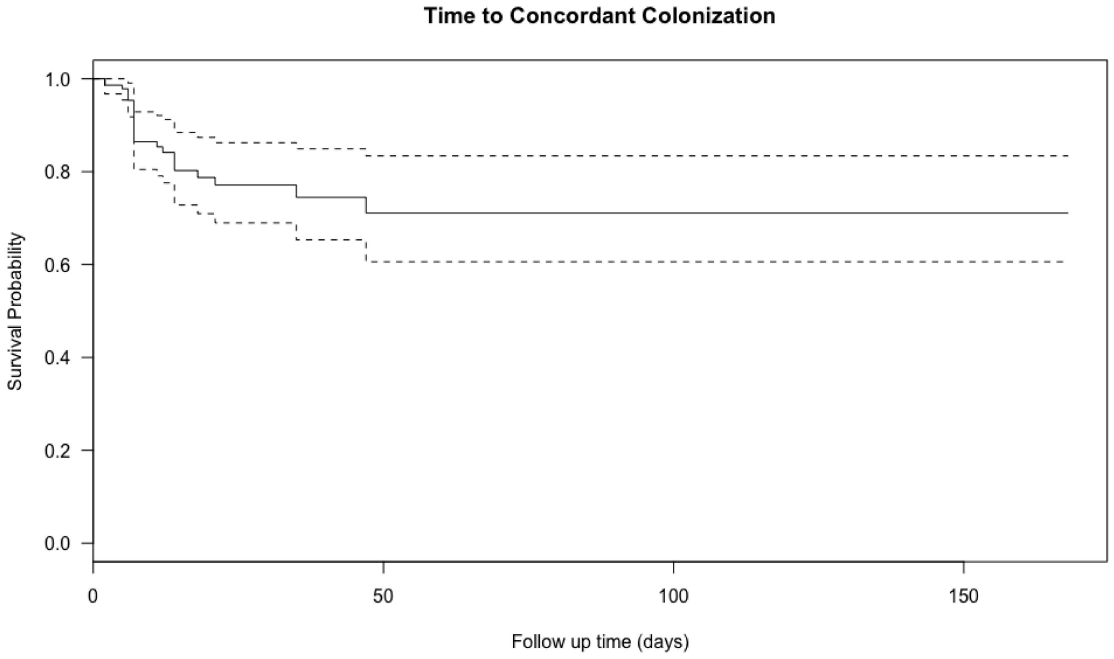
lower for inborn compared to outborn neonates (hazard ratio, HR: 0.65, 95% confidence interval, CI: 0.14 to 3.10). After conditioning on inborn and outborn status and birth weight, we estimate that neonates enrolled at the JHH site have 66% lower hazard of concordant colonization compared to those enrolled at the BMC site (HR: 0.34, 95% CI: 0.10 to 1.16). We also estimate that a 100 g increase in birth weight increases the hazard of concordant colonization by 2% after conditioning on admitted hospital and inborn and outborn status.

Table 3.2: Hazard Ratios from the Cox proportional hazards model adjusted for inborn and outborn status, admitted hospital, and birth weight.

Baseline Variable	Hazard Ratio	95% CI	P-value
Inborn vs Outborn	0.65	(0.14, 3.10)	0.59
JHH vs BMC	0.34	(0.10, 1.16)	0.08
Birth Weight (100g)	1.02	(0.97, 1.06)	0.41

Figures 3.1 and 3.2 display the estimated Kaplan-Meier and complementary log-log survival functions, respectively, for both concordant colonization and censoring (i.e. NICU discharge or death with administrative censoring at 90 days). The complementary log-log transformation of the survival curves are approximately linear with time (on the log scale), which is what would be expected if the underlying distribution of time to concordant colonization and censoring follow an exponential distribution, with mean 11.32 days and 26.44 days, respectively.

Figure 3.1: Kaplan-meier survival curve for time to concordant colonization (top) and time to censoring (bottom) over follow up time in days.



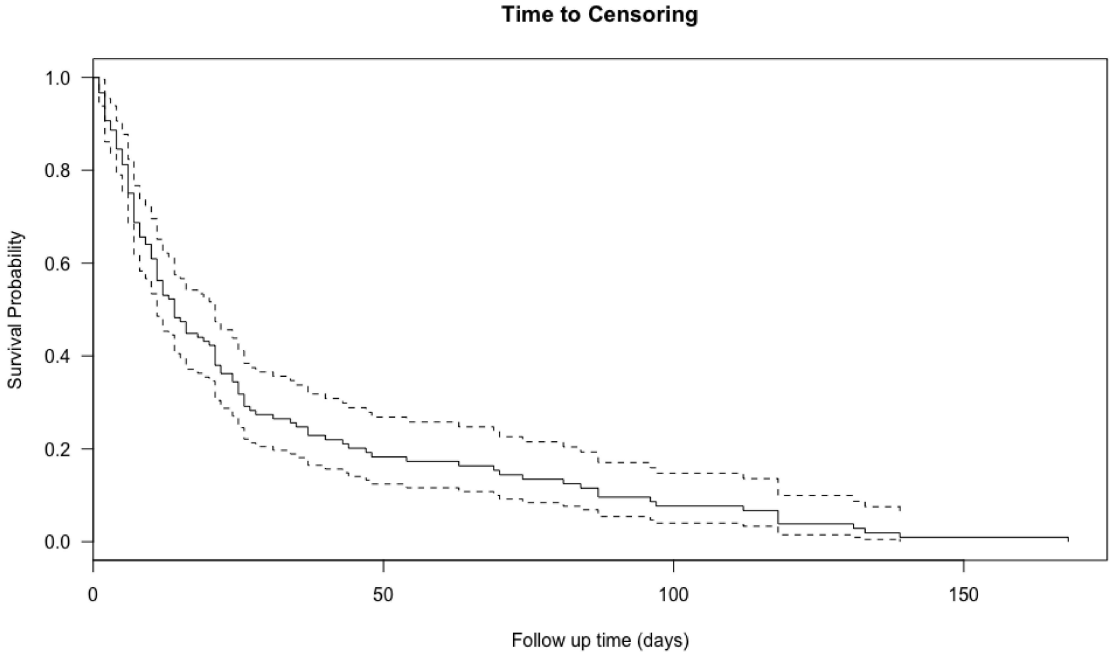
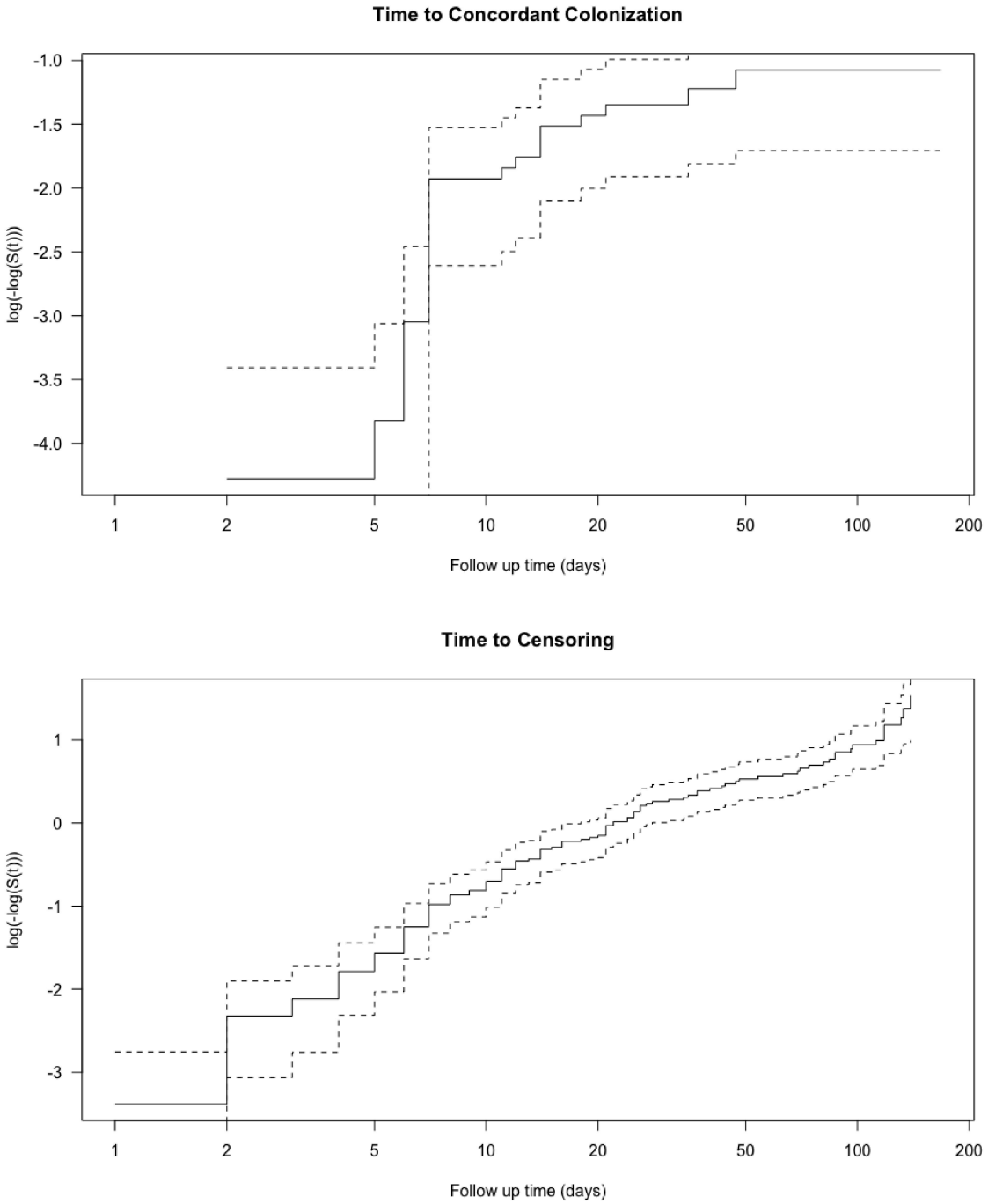


Figure 3.2: Complementary log-log plot of the estimated Kaplan Meier survival function for time to concordant colonization (top) and time to censoring (bottom) over follow up time in days.



3.3 Simulation study

We conducted a simulation study, based on the TREAT Parents trial, with the goal of evaluating the statistical performance of the adjusted estimators described in Sections 2.2 and 2.3 vs. the corresponding unadjusted estimator for the marginal treatment effects, i.e. log hazard ratio and risk difference, when the effective sample size of the RCT is small. We considered RCTs with effective sample sizes, i.e. the number of neonates who become concordantly colonized by 90 days, ranging from 20 to 100. We considered 3 scenarios which were defined by the degree to which the three baseline covariates are correlated with concordant colonization. In scenario 1, we assume that the baseline variables are independent of the outcome. In scenarios 2 and 3, we consider cases where the baseline variables are prognostic for the outcome, details provided below. In scenario 1, there is potential for efficiency loss for the adjusted estimators relative to the unadjusted estimator since the baseline variables are uncorrelated with the outcome, but asymptotically there may be no loss. In scenarios 2 and 3, there is potential for asymptotic efficiency gain when using the adjusted estimator relative to the unadjusted estimator, but it is not clear what will happen at finite sample sizes since the gains may be offset by losses due to added degrees of freedom in fitting the adjusted estimator. We separately consider the 3 scenarios under no and a benefit of the treatment.

For each scenario, 10,000 hypothetical trials were generated and the bias and variance of the estimated marginal treatment effects were computed, as well as, the

mean square error (MSE) and relative MSE (i.e. the ratio of unadjusted and adjusted estimator MSE).

3.3.1 Simulation design

In this section, we describe the data generating distributions for (W, A, T, Y) for each of the 3 scenarios and treatment effects, given the expected number of concordant colonizations $n = 20, 30, 40, 60$, and 100. W includes (X_1, X_2, X_3) corresponding to birth weight, participating hospital (JHH vs. BMC) and the indicator for inborn vs. outborn. For each n, we generate hypothetical trials of size $N = n \times 6$ based on the blinded data from the TREAT Parents trial where we observed 25 concordant colonizations in 150 recruited trial participants, i.e. a concordant colonization rate of 1/6.

The hypothetical trials of size N were generated as follows:

1. We took a sample of W of size N, with replacement, from the blinded data from the TREAT Parents trial. To avoid having a discrete population distribution of these baseline variables, an independent draw from the discrete uniform distribution on the range [-3,3] was added to the randomly selected birthweight (in grams).
2. Treatment was assigned independent of W by taking a draw from a Bernoulli distribution with probability 0.5 of being assigned to the treatment arm.
3. Motivated by the exploratory analysis of the blinded data from the TREAT

Parents trial, the time to concordant colonization (T^*) was set to follow an exponential distribution with hazard $\lambda = 1/17.32 \times exp(\gamma A + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$. The hazard of concordant colonization, λ, was computed for each neonate and then T^* was drawn from the respective exponential distribution. To generate distributions with no treatment benefit, γ was set to 0. To generate distributions with a treatment benefit, γ was set to -0.94 and -1.00 in scenarios 2 and 3, respectively; see note below for how we selected these values. We defined three scenarios based on the strength of the correlation between T^* and W as follows:

(a) For scenario 1, where T^* was uncorrelated with W, β_1 to β_3 were set equal to 0.

(b) For scenario 2, the correlation between T^* and W was set to the correlation observed in the blinded subset of the TREAT Parents trial, i.e. β_1 to β_3 were set equal to the estimated coefficients from Table 3.2.

(c) For scenario 3, the correlation between T^* and W was assumed to be stronger than that observed in the blinded subset of the TREAT Parents trial. Specifically, β_1 to β_3 were set equal to 1.5 times the estimated coefficients from Table 3.2.

Note: We generated outcome data using a proportional hazards model conditional on both A and W. Therefore, the assumption of proportion hazards for the marginal treatment effect does not necessarily hold. In our simulation study, the marginal proportional hazards assumption holds in the following cases: no

treatment effect (regardless of correlation between T^* and W) and benefit of treatment when T^* is uncorrelated with W. We next describe how we selected the values of γ for the simulation distributions corresponding to benefit of treatment. We did this by trial and error, where for each candidate value of γ we generated a very large number of simulated participants, administratively censored them at 90 days, and then fitted the Cox proportional hazards model with only a main term A; we tried different values of γ until the estimated coefficient on A was approximately equal to -0.92 (which is the marginal hazard ratio that the TREAT Parents trial was powered to detect). In simulated trials where the marginal proportional hazards model may be misspecified, the true value of the estimand was defined as the probability limit of the estimated coefficient on A in the marginal proportional hazards model, as sample size goes to infinity; we approximated this limit by fitting the Cox model to 1,000,000 independently generated data points with administrative censoring at 90 days.

4. An independent censoring time C was generated, assuming the distribution of censoring times followed an exponential distribution. The mean of the exponential distribution was calibrated to ensure that the targeted mean number of concordant colonizations was obtained, and varied by scenario and treatment effect.

   (a) For scenario 1, 2.27 and 3.33 days for no and positive treatment benefit

   (b) For scenario 2, 2.17 and 3.23 days for no and positive treatment benefit

(c) For scenario 3, 2.04 and 3.13 days for no and positive treatment benefit

5. Lastly, given the values of T^* and C, T and Y were computed.

3.3.2 Estimators

For each hypothetical trial generated in the simulation study, we fit the adjusted estimators described in Sections 2.2 and 2.3 and the corresponding unadjusted estimators for the marginal treatment effects, i.e. log hazard ratio and risk difference. In this section, we provide details for implementation of the adjusted estimators.

For the primary endpoint, i.e. time to concordant colonization, the unadjusted marginal treatment effect is obtained by fitting a Cox proportional hazards model with a main term A. To obtain the adjusted marginal treatment effect, the method of Lu and Tsiatis was implemented including main terms for X_1, X_2, and X_3 in both the linear models for baseline hazards and hazard for the censoring distribution. Note that in scenario 1, both of these linear models contain the true model; however, in scenarios 2 and 3, these linear models are not correctly specified.

For the secondary endpoint, i.e. the binary indicator of concordant colonization by 90 days after randomization, the marginal treatment effect is the risk difference and the unadjusted estimator is the difference between the proportion of neonates concordantly colonized in the treatment and placebo arms. Three marginal adjusted estimators were implemented, and these include the model standardization estimator, the DR-WLS estimator and PLEASE. All three of the adjusted estimators require

specification of a working outcome regression model; which was defined as a logistic regression model for the binary indicator of concordant colonization by 90 days and includes main terms for X_1, X_2, and X_3. The DR-WLS estimator and PLEASE require an additional working model, the propensity score model for treatment assignment. We defined the propensity score model as a logistic model for A as a function of main terms for X_1, X_2, and X_3. The working propensity score model is correctly specified as the true model, i.e. a logistic model for treatment assignment that contains only an intercept, is contained within our working model. The working outcome regression model is not required to be correctly specified in order to obtain consistent estimates of the marginal treatment effect; the doubly robust properties of the DR-WLS estimator and PLEASE guarantee consistency of the estimators when the working propensity score model is correctly specified.

Chapter 4

Results

We conducted simulation studies considering effective sample sizes of 20, 30, 40, 60 and 100, thus generating hypothetical trials of size 120, 180, 240, 360, and 600 subjects, respectively. In this section, we provide the results of our simulation studies for both the time to event outcome (primary endpoint) and binary outcome (secondary endpoint). For each outcome, we evaluated the performance of the unadjusted and adjusted estimators by computing the bias, variance, and mean squared error (MSE). In addition, we computed the relative MSE defined as the MSE of the unadjusted estimator divided by the MSE of the adjusted estimator.

4.1 Primary Endpoint

Table 4.1 summarizes the results of the simulation study when there is no benefit of the treatment. The bias of both the unadjusted and adjusted estimators is small for all scenarios and effective sample sizes. As expected, within a scenario (i.e. data generated under the same assumed correlation between T^* and W), we observe decreasing variance with increasing effective sample size. For scenario 1, where T^* and W are uncorrelated, there is a loss of precision for estimating the marginal treatment effect when using the adjusted estimator compared to the unadjusted estimator. For instance, the precision loss is roughly 5% for trials with effective sample size of 20. This precision loss decreases as the effective sample size increases; however, we note that even at an effective sample size of 60, the precision loss is 2.3%. In scenario 2, where the correlation between T^* and W is the correlation observed in the blinded subset of the TREAT Parents trial data, we observe a 3 and 1% loss of precision for effective sample sizes of 20 and 30, respectively, when using the adjusted estimator compared to the unadjusted estimator. However, for effective sample sizes of 40 and greater, we observe precision gains on the order of 1 to roughly 2% when using the adjusted estimator relative to the unadjusted estimator. In scenario 3, where the correlation between T^* and W is increased by 50% compared to what we observed in the blinded subset of the TREAT Parents trial data, the precision gains are observed starting at effective sample sizes of 30 and range from 3% to 8%.

Table 4.2 shows the results from the simulation study where we assume a benefit

Table 4.1: Results of the simulation study assuming no treatment effect. For each scenario and effective sample size (n), the results summarize 10,000 hypothetical trials. Bias, variance and MSE are computed for the unadjusted (labeled as "Unadj") and adjusted (labeled as "Adj") estimators for the marginal log hazard ratio. The relative MSE compared the unadjusted to adjusted estimator.

		Bias		Variance		MSE		Relative MSE
Scenario	n	Unadj	Adj	Unadj	Adj	Unadj	Adj	Unadj/Adj
	20	0.0042	0.0048	0.2423	0.2539	0.2423	0.2539	0.954
	30	-0.0002	-0.0006	0.1497	0.1549	0.1497	0.1549	0.966
1	40	-0.0007	-0.0016	0.1076	0.1116	0.1076	0.1116	0.964
	60	-0.0018	-0.0015	0.0698	0.0715	0.0699	0.0715	0.977
	100	-0.0029	-0.0037	0.0405	0.0411	0.0405	0.0411	0.985
	20	0.0017	0.0000	0.2370	0.2449	0.2370	0.2449	0.968
	30	-0.0008	-0.0013	0.1411	0.1424	0.1411	0.1424	0.991
2	40	-0.0038	-0.0025	0.1089	0.1083	0.1089	0.1083	1.005
	60	-0.0070	-0.0074	0.0687	0.0672	0.0687	0.0673	1.021
	100	0.0000	0.0001	0.0406	0.0397	0.0406	0.0397	1.023
	20	-0.0051	-0.0057	0.2396	0.2416	0.2396	0.2416	0.992
	30	-0.0087	-0.0080	0.1481	0.1443	0.1482	0.1444	1.026
3	40	0.0001	-0.0013	0.1099	0.1049	0.1099	0.1049	1.048
	60	0.0011	0.0018	0.0712	0.0668	0.0712	0.0668	1.066
	100	-0.0019	-0.0024	0.0424	0.0391	0.0424	0.0391	1.084

of the treatment, such that the marginal log hazard ratio is -0.92. For all scenarios, we observe bias in the estimate of the marginal treatment effect for effective sample sizes of 20 to 40; with greater bias in the unadjusted compared to the adjusted estimator. Otherwise, the general patterns we observed for the no treatment effect simulation setting are replicated when assuming a treatment benefit. If the joint distribution of (T, Y, W) in the completed TREAT Parents trial is similar to the distribution we have assumed in our simulation study, then any gains from adjustment would be negligible (at 40 concordant colonizations, precision gains of 0.34%).

In the simulation study, we used the *speffSurv* function, see R package 'sp-

Table 4.2: Results of the simulation study assuming a benefit of the treatment. For each scenario and effective sample size (n), the results summarize 10,000 hypothetical trials. Bias, variance and MSE are computed for the unadjusted (labeled as "Unadj") and adjusted (labeled as "Adj") estimators for the marginal log hazard ratio. The relative MSE compared the unadjusted to adjusted estimator.

		Bias		Variance		MSE		Relative MSE
Scenario	n	Unadj	Adj	Unadj	Adj	Unadj	Adj	Unadj/Adj
	20	-0.0438	-0.0295	0.2983	0.3175	0.3002	0.3184	0.943
	30	-0.0311	-0.0175	0.1843	0.1934	0.1852	0.1937	0.957
1	40	-0.0190	-0.0046	0.1337	0.1372	0.1341	0.1372	0.977
	60	-0.0087	0.0043	0.0841	0.0860	0.0842	0.0860	0.979
	100	-0.0051	0.0058	0.0490	0.0502	0.0491	0.0502	0.977
	20	-0.0491	-0.0340	0.2880	0.3076	0.2905	0.3087	0.941
	30	-0.0357	-0.0191	0.1859	0.1909	0.1872	0.1913	0.979
2	40	-0.0250	-0.0069	0.1308	0.1309	0.1314	0.1310	1.003
	60	-0.0127	0.0065	0.0833	0.0825	0.0835	0.0825	1.012
	100	-0.0072	0.0087	0.0490	0.0479	0.0491	0.0480	1.023
	20	-0.0479	-0.0347	0.2899	0.3064	0.2922	0.3076	0.950
	30	-0.0312	-0.0144	0.1878	0.1834	0.1888	0.1836	1.028
3	40	-0.0204	-0.0024	0.1300	0.1237	0.1305	0.1237	1.055
	60	-0.0154	0.0034	0.0833	0.0782	0.0835	0.0782	1.068
	100	-0.0093	0.0083	0.0490	0.0455	0.0491	0.0456	1.078

eff2trial' [17], to estimate the "adjusted" marginal log hazard ratio, the marginal log hazard ratio estimated by leveraging prognostic baseline covariates. For small effective sample sizes, the function sometimes did not find a solution for the "adjusted" log hazard ratio due to convergence problem; approximately 3 and 24 out of 10,000 hypothetical trials, on average, at an effective sample size of 20 when we assumed no and a benefit of the treatment, respectively, and 2 hypothetical trials at an effective sample size of 30 when we assumed a benefit of the treatment.

4.2 Secondary Endpoint

We now summarize the simulation results for the secondary outcome, the binary indicator for concordant colonization within 90 days of randomization. Recall the marginal treatment effect of interest is the risk difference. Table 4.3 displays the results from the simulation study assuming that the treatment has no effect. For all scenarios, the bias is small and similar regardless of the effective sample size. In scenario 1, where the Y and W are uncorrelated, there are losses in precision for all the adjusted estimators relative to the unadjusted estimator, as expected. The loss in precision decreases with increasing effective sample size and we note that the PLEASE has greater losses in precision compared to the standardization and DR-WLS estimators. In scenario 2, where the correlation between Y and W is based on the observed correlation within the blinded subset of the TREAT Parents trial data, there are precision losses for effective sample sizes 20 and 30. For effective sample sizes 40 and above, we see a roughly 1 to 2% improvement in precision to estimate the marginal risk difference comparing the standardization and DR-WLS estimators to the unadjusted estimators. At effective sample sizes 40 and 60, the PLEASE demonstrates precision losses, but for effective sample size 100, it demonstrates a precision gain which is slightly smaller than the precision gains we observe for the standardization and DR-WLS estimators. When we inflate the observed correlation between Y and W, in scenario 3, we see precision gains on the order of 1 to 6% for effective sample sizes of 30 and greater, with slightly smaller precision gains for the

PLEASE.

Table 4.4 displays the results of the simulation study where we assumed a benefit of the treatment. Here the marginal treatment effect reduces the risk of concordant colonization within 90 days by 12.5%, absolute risk reduction. Overall, we observed similar patterns of results compared to when we assumed no treatment effect (see Table 4.3).

4.3 Illustrative example

In this section, we use the following data generating distributions for (W, A, T, Y) as another illustrative example to demonstrate the relationship between the number of events and relative efficiency by varying degrees of correlation between the time to event endpoint and baseline covariates. We consider only one baseline covariate $W = X_1$. For each targeted number of events, $n = 10, 20, \cdots, 80$, we generate 10,000 hypothetical trials of size $N = n \times 6$, given the degree of correlation varied from 0 to 0.5 at an increment of 0.1, i.e $\beta = 0, 0.1, 0.2, 0.3, 0.4$ and 0.5.

The hypothetical trials of size N are generated in a similar manner as follows: We generated X_1, a random sample of size N from a Normal distribution with mean 0 and standard deviation 1. Treatment, A, was assigned by taking a draw from a Bernoulli distribution with probability 0.5 of being assigned to the treatment arm. The time to event, T^*, assumed to follow an exponential distribution with mean 17.32 days, was

generated with hazard $\lambda = 1/17.32 \times exp(\gamma A + \beta X_1)$, where β indicates the degree of correlation. An independent time to censoring, C, was generated, assuming the distribution of censoring time followed an exponential distribution with mean 2.27 days when $\beta = 0$ and 0.1, 2.22 days when $\beta = 0.2$ and 0.3, and 2.13 days when $\beta = 0.4$ and 0.5. Lastly, given T^* and C, T and Y were computed. Since we observed similar patterns from the simulation results when we assumed no and a benefit of the treatment, we only consider the no treatment effect simulation setting where $\gamma = 0$.

Figure 4.1: Relationship between relative MSE and effective sample size by the degree of correlation between the time to event outcome and baseline covariate assuming no treatment effect.

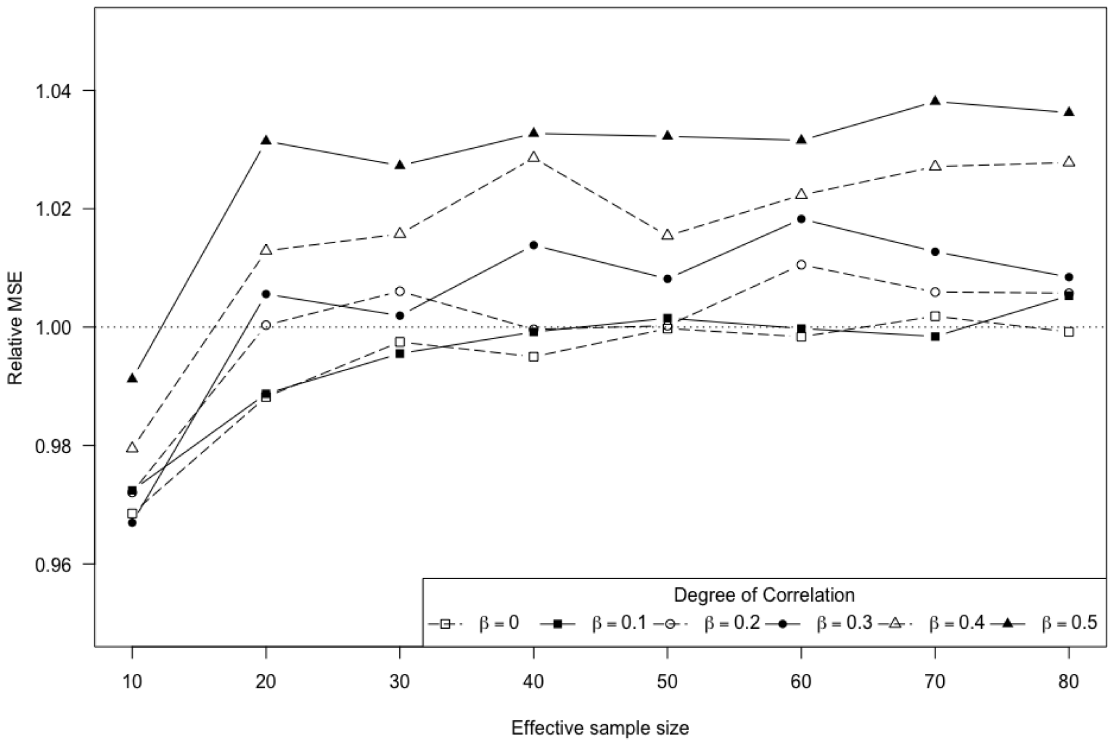


Figure 4.1 shows the results of the simple simulation study described above. It

shows that the potential precision gains in the estimated treatment effect depend on the strength of the correlation between the outcome and baseline covariate and increase with increasing effective sample size. When the outcome and baseline covariate are uncorrelated, there is a loss of precision, but this precision loss decreases as the effective sample size increases. When the outcome and baseline covariate are correlated, there is no precision gain at an effective sample size of 10; however, for effective sample sizes of 20 and greater, we start to observe precision gains for the degrees of correlation of 0.2 and greater. In summary, we observe that the higher the correlation between the outcome and baseline covariate, the higher the precision gain in the estimated treatment effect, given a fixed effective sample size.

Table 4.3: Results of the simulation study assuming no treatment effect. For each scenario and effective sample size (n), the results summarize 10,000 hypothetical trials. Bias, variance and MSE are computed for the unadjusted (labeled as "RD") and adjusted (labeled as "ST" for the standardization estimator, "DR" for the doubly-robust estimator, "PL" for the PLEASE) estimators for the marginal risk difference. The relative MSE compared the unadjusted to adjusted estimator.

		Bias				Variance				MSE				Relative MSE		
Scenario	n	RD	ST	DR	PL	RD	ST	DR	PL	RD	ST	DR	PL	RD/ST	RD/DR	RD/PL
	20	0.0007	0.0007	0.0008	0.0009	0.0047	0.0049	0.0049	0.0049	0.0047	0.0049	0.0049	0.0049	0.975	0.974	0.959
	30	0.0006	0.0007	0.0007	0.0007	0.0032	0.0032	0.0032	0.0033	0.0032	0.0032	0.0032	0.0033	0.981	0.980	0.965
1	40	0.0004	0.0003	0.0003	0.0003	0.0023	0.0024	0.0024	0.0024	0.0023	0.0024	0.0024	0.0024	0.987	0.987	0.975
	60	-0.0002	-0.0002	-0.0002	-0.0002	0.0015	0.0016	0.0016	0.0016	0.0015	0.0016	0.0016	0.0016	0.990	0.989	0.982
	100	0.0003	0.0003	0.0003	0.0004	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.994	0.994	0.990
	20	0.0035	0.0037	0.0037	0.0037	0.0047	0.0047	0.0047	0.0048	0.0047	0.0047	0.0047	0.0048	0.995	0.994	0.971
	30	0.0031	0.0029	0.0029	0.0030	0.0030	0.0030	0.0030	0.0031	0.0030	0.0030	0.0030	0.0031	0.997	0.997	0.979
2	40	0.0019	0.0019	0.0019	0.0020	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	1.013	1.012	0.998
	60	0.0029	0.0028	0.0028	0.0028	0.0016	0.0015	0.0015	0.0016	0.0016	0.0015	0.0015	0.0016	1.009	1.008	0.999
	100	0.0024	0.0024	0.0024	0.0024	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	1.020	1.020	1.016
	20	-0.0025	-0.0023	-0.0023	-0.0022	0.0047	0.0046	0.0046	0.0047	0.0047	0.0046	0.0046	0.0047	1.016	1.015	0.990
	30	-0.0026	-0.0026	-0.0026	-0.0025	0.0031	0.0030	0.0030	0.0030	0.0031	0.0030	0.0030	0.0031	1.030	1.029	1.008
3	40	-0.0018	-0.0020	-0.0020	-0.0019	0.0024	0.0023	0.0023	0.0023	0.0024	0.0023	0.0023	0.0023	1.047	1.046	1.033
	60	-0.0014	-0.0013	-0.0013	-0.0013	0.0016	0.0015	0.0015	0.0015	0.0016	0.0015	0.0015	0.0015	1.049	1.048	1.039
	100	-0.0020	-0.0021	-0.0021	-0.0020	0.0010	0.0009	0.0009	0.0009	0.0010	0.0009	0.0009	0.0009	1.061	1.061	1.057

Table 4.4: Results of the simulation study assuming a benefit of the treatment. For each scenario and effective sample size (n), the results summarize 10,000 hypothetical trials. Bias, variance and MSE are computed for the unadjusted (labeled as "RD") and adjusted (labeled as "ST" for the standardization estimator, "DR" for the doubly-robust estimator, "PL" for the PLEASE) estimators for the marginal risk difference. The relative MSE compared the unadjusted to adjusted estimator.

		Bias				Variance				MSE				Relative MSE		
Scenario	n	RD	ST	DR	PL	RD	ST	DR	PL	RD	ST	DR	PL	RD/ST	RD/DR	RD/PL
	20	0.0029	0.0030	0.0030	0.0031	0.0046	0.0047	0.0047	0.0048	0.0046	0.0047	0.0047	0.0048	0.969	0.968	0.950
	30	0.0020	0.0020	0.0020	0.0019	0.0030	0.0031	0.0031	0.0031	0.0030	0.0031	0.0031	0.0031	0.981	0.981	0.969
1	40	0.0021	0.0020	0.0020	0.0021	0.0023	0.0023	0.0023	0.0024	0.0023	0.0023	0.0023	0.0024	0.986	0.986	0.977
	60	0.0019	0.0019	0.0019	0.0019	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.994	0.993	0.983
	100	0.0020	0.0020	0.0020	0.0019	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.997	0.996	0.993
	20	-0.0014	-0.0012	-0.0012	-0.0010	0.0045	0.0046	0.0046	0.0047	0.0045	0.0046	0.0046	0.0047	0.982	0.979	0.957
	30	-0.0016	-0.0016	-0.0016	-0.0018	0.0031	0.0030	0.0031	0.0031	0.0031	0.0031	0.0031	0.0031	1.002	1.002	0.987
2	40	-0.0019	-0.0019	-0.0019	-0.0018	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	1.011	1.011	0.993
	60	-0.0005	-0.0005	-0.0005	-0.0005	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	1.010	1.009	0.996
	100	-0.0010	-0.0010	-0.0010	-0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	1.019	1.019	1.015
	20	-0.0019	-0.0018	-0.0018	-0.0019	0.0044	0.0044	0.0044	0.0045	0.0044	0.0044	0.0044	0.0045	1.016	1.013	0.983
	30	-0.0019	-0.0018	-0.0018	-0.0018	0.0031	0.0030	0.0030	0.0030	0.0031	0.0030	0.0030	0.0031	1.047	1.046	1.028
3	40	-0.0014	-0.0015	-0.0015	-0.0016	0.0022	0.0021	0.0021	0.0022	0.0022	0.0021	0.0021	0.0022	1.051	1.051	1.034
	60	-0.0021	-0.0022	-0.0022	-0.0022	0.0015	0.0014	0.0014	0.0014	0.0015	0.0014	0.0014	0.0014	1.054	1.054	1.045
	100	-0.0024	-0.0024	-0.0024	-0.0024	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	1.058	1.058	1.051

Chapter 5

Discussion

In this work, we have evaluated the performance of several adjusted estimators of marginal treatment effects within RCTs with small sample sizes. Small sample size RCTs may include pilot RCTs evaluating feasibility to recruit, retain and treat patients, Phase II regulatory trials, confirmatory RCTs of expensive interventions or confirmatory RCTs conducted for rare outcomes. We were motivated by the TREAT Parents trial, being conducted in two Johns Hopkins Hospital system NICUs, evaluating the impact of treating parents colonized with *Staph aureus* to minimize the transmission of the parental strain of *Staph aureus* to the hospitalized neonate. This study required a total of 40 concordantly colonized neonates to detect the hypothesized benefit of treating the parents with nasal antibiotics and chlorhexidine bathing compared to placebo (log hazard ratio of -0.92) with 80% power. We evaluated the potential benefits of using baseline covariate adjustment on the primary (time to con-

cordant colonization) and secondary (ever concordantly colonized by 90 days post randomization) endpoints in the TREAT Parents study.

We designed simulation studies that reflected the observed characteristics within a blinded subset of the TREAT Parents trial data. Based on the targeted effective sample size for the trial ($n = 40$) and hypothesized treatment effect, we found that the adjusted estimators for both the log hazard ratio for the time to event endpoint and risk difference for the binary endpoint would result in no substantial gain in precision compared to using the corresponding unadjusted estimators. However, in cases where the baseline covariates demonstrate greater prognostic ability for the primary and secondary endpoints (e.g. increased correlation by 50%), we observed precision gains of roughly 5% at effective sample sizes as small as 30.

For researchers designing RCTs with small effective sample sizes, the decision to use an adjusted estimator for the marginal treatment effect should be driven by the degree to which a small number of *a priori* selected baseline variables are correlated with the outcome of interest. Our simulation studies demonstrate that for a fixed effective sample size (e.g. $n = 30$), the potential precision loss if the baseline variables are uncorrelated or only weakly correlated with the primary outcome (roughly 2%) can be roughly the same as the potential precision gain if the baseline variables are more strongly correlated with the primary outcome. Therefore, researchers planning such RCTs should take care to evaluate the potential prognostic ability of the selected baseline variables.

If possible, we recommend evaluating whether baseline variables are prognostic for the outcome of interest using data from observational studies or completed RCTs of similar patient populations. For the time to event outcome, we summarized the strength of association using the Cox proportional hazards model. For outcomes that follow an exponential family distribution, the relative R-square calculation proposed by Moore and van der Laan [14] may be applied to existing data. Alternatively, simulation studies, as demonstrated in Chapter 3 and Chapter 4 (see Section 4.3) may be used to determine if, under the specified assumptions, there is likely a benefit of using baseline variable adjusted estimators of the marginal treatment effect.

Our work has several limitations that should be addressed moving forward. First, we have designed our simulation studies using parametric survival models. It would be useful to consider additional simulation settings with different parametric survival models or design simulation settings based on resampling data from completed trials to generate hypothetical trials from unknown data generating distributions. Second, we explored the performance of the adjusted estimator of Lu and Tsiatis based on data generating distributions where censoring was independent of treatment and W. It would be informative to evaluate the performance of this method under violations of this assumption. Third, there are alternative adjusted estimators that have been proposed for time to event outcomes, see [6]. Diaz et al propose an adjusted estimator for the difference in mean restricted survival time to day τ comparing treatment to placebo arms that, asymptotically, is guaranteed to be as precise or more precise

than the unadjusted estimator and can account for censoring that may depend on the treatment assignment or baseline variables. It would be useful to evaluate the performance of this estimator under small effective sample sizes. Lastly, we used the coefficients from the Cox proportional hazards model as a way to measure correlation between the baseline variables and the time to event outcome. It would be useful to explore the potential for a statistic, similar to the relative R-square cited above, that could be used as a guide for the prognostic ability of baseline variables for time to event outcomes.

Bibliography

[1] S. F. Assmann, S. J. Pocock, L. E. Enos, and L. E. Kasten, "Subgroup analysis and other (mis)uses of baseline data in clinical trials," *The Lancet*, vol. 355, pp. 1064–1069, 2000.

[2] S. J. Pocock, S. E. Assmann, L. E. Enos, and L. E. Kasten, "Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems," *Statistics in Medicine*, vol. 21, pp. 2917–2930, 2002.

[3] E. Colantuoni and M. Rosenblum, "Leveraging prognostic baseline variables to gain precision in randomized trials," *Statistics in Medicine*, vol. 34, pp. 2602–2617, 2015.

[4] S. Gruber and M. J. van der Lann, "Targeted minimum loss based estimator that outperforms a given estimator," *The International Journal of Biostatistics*, vol. 8, no. 1, p. Article 11, 2012.

[5] Z. Tan, "Bounded, efficient and doubly robust estimating equations for marginal and nested structural models," *Biometrika*, vol. 97, pp. 661–682, 2010.

[6] I. Díaz, E. Colantuoni, D. F. Hanley, and M. Rosenblum, "Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards," *Lifetime Data Analysis*, 2018.

[7] X. Lu and A. A. Tsiatis, "Improving the efficiency of the log-rank test using auxiliary covariates," *Biometrika*, vol. 95, no. 3, pp. 679–694, 2008.

[8] I. Díaz, E. Colantuoni, and M. Rosenblum, "Enhanced precision in the analysis of randomized trials with ordinal outcomes," *Biometrics*, vol. 72, no. 2, pp. 422–431, 2016.

[9] J. A. Steingrimsson, D. F. Hanley, and M. Rosenblum, "Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions," *Contemporary Clinical Trials*, vol. 54, pp. 18–24, 2017.

[10] D. F. Hanley, K. Lane, N. McBee, W. Ziai, S. Tuhrim, K. R. Lees, J. Dawson, D. Gandhi, N. Ullman, A. Mould, S. W. Mayo, A. D. Mendelow, B. Gregson, K. Butcher, P. Vespa, D. W. Wright, C. S. Kase, J. R. Carhuapoma, P. M. Keyl, M. Diener-West, J. Muschelli, J. F. Betz, C. B. Thompson, E. A. Sugar, G. Yenokyan, S. Janis, S. John, S. Harnof, G. A. Lopez, E. F. Aldrich, M. R. Harrigan, S. Ansari, J. Jallo, J.-L. Caron, D. LeDoux, O. Adeoye, M. Zuccarello, H. P. Adams Jr., M. Rosenblum, R. E. Thompson, and I. A. A. for the CLEAR III Investigators, "Thrombolytic removal of intraventricular haemor-

rhage in treatment of severe stroke: results of the randomised, multicentre, multiregion, placebo-controlled CLEAR III trial," *The Lancet*, vol. 389, pp. 603–611, 2017.

[11] W. A. Mould, J. R. Carhuapoma, J. Muschelli, K. Lane, T. C. Morgan, N. A. McBee, A. J. Bistranhall, N. L. Ullman, P. Vespa, N. A. Martin, I. Awad, M. Zuccarello, and D. F. Hanley, "Minimally invasive surgery plus rt-pa for intracerebral hemorrhage evacuation (MISTIE) decreases perihematomal edema," *Stroke*, vol. 44, no. 3, pp. 627–634, 2013.

[12] A. M. Milstone, D. W. Koontz, A. Voskertchian, V. O. Popoola, K. Harrelson, T. Ross, S. W. Aucott, M. M. Gilmore, K. C. Carroll, and E. Colantuoni, "Treating parents to reduce NICU transmission of *Staphylococcus aureus* (TREAT PARENTS) trial: protocol of a multisite randomized, double-blind, placebo-controlled trial," *BMJ Open*, vol. 5, p. e009274, 2015.

[13] D. R. Cox, "Partial likelihood," *Biometrika*, vol. 62, no. 2, pp. 269–276, 1975.

[14] K. Moore and M. J. van der Lann, "Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation," *Statistics in Medicine*, vol. 28, no. 1, pp. 39–64, 2009.

[15] J. Robin, M. Sued, Q. Lei-Gomez, and A. Rotnitzky, "Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable," *Statistical Science*, vol. 22, no. 4, pp. 544–559, 2007.

[16] A. Rotnitzky, Q. Lei, M. Sued, and J. M. Robins, "Improved double-robust estimation in missing data and causal inference models," *Biometrika*, vol. 99, no. 2, pp. 439–456, 2012.

[17] M. Juraska, with contributions from Peter B. Gilbert, X. Lu, M. Zhang, M. Davidian, and A. A. Tsiatis, *speff2trial: Semiparametric efficient estimation for a two-sample treatment effect*, 2012, r package version 1.0.4. [Online]. Available: https://CRAN.R-project.org/package=speff2trial

Vita

Su Jin Lim was born on May 1, 1992 in Seoul, South Korea. She received her Bachelor of Arts degree in Mathematics from Grinnell College in May 2016, where she graduated as a Phi Beta Kappa. In September 2016, she enrolled in the Master of Science (ScM) Biostatistics program at the Johns Hopkins Bloomberg School of Public Health.